<LECTURE>Trusting Digital Records: the Major Findings of the InterPARES Project

AUTHOR(S):

DURANTI, Luciana

CITATION:

DURANTI, Luciana. <LECTURE>Trusting Digital Records: the Major Findings of the InterPARES Project. 京都大学大学文書館研究紀要 2013, 11: 15-33

ISSUE DATE:

2013-03-21

URL:

https://doi.org/10.14989/173414

講　演

Trusting Digital Records:

the Major Findings of the InterPARES Project

Luciana Duranti †
Translated by Koga Takashi ‡

The Goal of InterPARES 1 and 2 (1998-2006)

InterPARES began in 1998 with the purpose of developing the body of theory and methods necessary to ensure that digital records produced in databases and office systems, as well as in dynamic, experiential, and interactive systems in the course of artistic, scientific, and e-government activities can be created in an accurate and reliable form, and maintained and preserved in an authentic form, both in the long and short term, for the use of those who created them and for society at large, regardless of technological obsolescence and media fragility.

In other words, InterPARES research was meant to develop new theory and new methodology for digital preservation, based on the understanding that preservation begins at creation.

Goal of InterPARES 3 (2007-2012)

In contrast, InterPARES 3 had the purpose of applying the theory and methodology developed during the initial two phases of the project towards the solution of existing problems, especially in institutions and organizations with few resources.

Key InterPARES 1 & 2 Products

The key products of InterPARES 1 and 2 were tested in each organization acting as a test-bed and adapted as needed. All these key products, plus the other products that are available on the InterPARES website are not covered by copyright. This means that anyone can download them, copy them, translate them, distribute them, reissue them, and adapt them to specific needs.

The first key product is the "Policy Framework," which is a framework of principles for guiding the development of policies for records creating and preserving organizations. The Policy Framework comprises thirteen principles for records creators, and thirteen principles for records preservers.

The second key product is the "Creator Guidelines." They include recommendations for making and maintaining digital materials, directed to individuals, professionals, and small communities of practice, for example, the doctor's office, the law office, and other organizations that are too small to afford a records manager. These

† Professor, School of Library, Archival and Information Studies, The University of British Columbia
‡ Associate Professor, Faculty of Human Studies, Tenri University

guidelines guide any person, office or organization who uses them to create digital records so that can be maintained in the right way, and to keep and use them ensuring that their accuracy, reliability and authenticity are protected and that they remain accessible through time.

These guidelines have already been translated into several languages. On the website they are posted in English, French, Spanish, Catalan, Portuguese, and Chinese. They have not been translated in Japanese yet, but, if useful, the Japanese professional community is very welcome to do so.

The third key product is the "Preserver Guidelines" : recommendations for digital preservation for archival institutions, programs, units, and organizations. They are basically the complement to the Creator Guidelines as they build on them. These are for archivists or whomever is in charge of preservation. They represent all the digital guidance that is needed to preserve digital records received from the creator. These guidelines have been translated into the same languages as the Creator Guidelines.

The fourth key product is the "Benchmarks and Baseline Requirements for Authenticity." Benchmarks requirements for authenticity are the requirements for those who create and maintain the records, to make sure that the records can be proven to be authentic at any given time in their active life. Baseline requirements are the requirements for archivists or any preserver to maintain authenticity over the long term and to be able to demonstrate it.

The fifth key product is the "File Format Selection Guidelines," which articulate principles and criteria for selecting the file formats, wrappers or encoding schemes that are the most appropriate for preservation.

The sixth key product is the "Terminology Database," which is composed of three parts: a glossary, a dictionary, and three ontologies. The glossary comprises the terms contained in InterPARES documents, and defines them as used by the InterPARES researchers. The dictionary includes the same terms, but defined in several other ways, using as sources existing glossaries and dictionaries, also from other fields. The three ontologies are graphics which show the relationships among terms and concepts.

The seventh key product is constituted of two records management models. These are extremely important for those who want to design systems or analyze what they have, identifying possible gaps. The "Chain of Preservation Model," or COP model, follows the concept of the lifecycle of records from creation to preservation. The "Business-driven Recordkeeping Model," or BRM, in contrast to the COP model, follows the Australian concept of the continuum. If you believe in that concept, you can use that specific model to design systems that enact that concept. One might guess from what I will discuss in the course of this presentation that I do not support the BRM. I do support the COP model. But InterPARES wishes to serve the records professional community worldwide, thus, it developed models that can accommodate all points of view.

The eighth and ninth products are two books which are available online. The first resulted from the InterPARES 1 project, and the second from InterPARES 2. [1]

InterPARES 3 International Alliance

The InterPARES 3 International Alliance comprised the following teams representing countries or regions: Canada (including United States partners); Brazil; Catalonia (Spain); China; Colombia; Italy; Korea; Malaysia; Mexico; Norway (till 2009); Singapore (till 2009); and Turkey. The project was funded by the Social Sciences

and Humanities Research Council of Canada and the University of British Columbia(UBC) in Vancouver, Canada. The project headquarters resided in the School of Library, Archival and Information Studies, at UBC. I was the Director.

InterPARES 3 Findings

The findings of InterPARES 3 were of three types: conceptual findings (I will discuss, among these, the concept of trustworthiness), methodological findings (I will discuss, among these, preservation methods), and strategic findings (I will discuss, among these, the role of the archivist).

Digital vs. Traditional Records

In the digital environment, the content, the structure or form, and the medium of the records are no longer linked to each other. They exist separately.

The entity "record" that is stored in the system is separate from what one can see on the computer monitor. In the system, we have content data, form data, and composition data, but what we see as the manifestation of those data on the screen is a different thing: it is the documentary view of the digital components of the record.

When we save a record, what we do is to break it down into its digital components, and when we recall the record, what we do is to put those components together again, creating a copy. Accessing a stored record involves making a reproduction. This means that we cannot preserve a digital record; we can only preserve our capacity of re-producing it.

This is a vital difference between traditional records and digital records. We can never test the authenticity of digital records on the records themselves, because a digital record is made up of many parts, those inside the system and what we see on the screen, and also because what we retrieve at any given time is always a new reproduction of what we saved. With traditional records, to test authenticity we could analyze the paper, the ink, the seal of the record as it originally existed, whether its status of transmission were that of original, draft, or copy.[2] We could see whether the record was materially the thing it claimed to be or not. Digital records are always new and, in order to have substantial evidence of their authenticity, we must infer their trustworthiness from the environment in which they exist and a documented chain of custody.

In other words, our assessment of trustworthiness with digital records is entirely based on how much we trust the person(s) and the systems making and holding them overtime. As records professionals we do say that we believe in what the context will tell us. This is the time to test the truth of such statement.

Trust & Its Rules

What is trust? Trust involves acting without the specific knowledge that we need to act. Thus, it consists of substituting the information that we do not have with other information. For example, I do not understand Japanese. So I do not know what the translator of my presentation is telling the audience. However, I know that Professor Masahito Ando thinks that he can translate my presentation accurately. I know Professor Ando and I believe that his opinion is accurate. Thus, I use my knowledge of Professor Ando to make up for the knowledge I don't have of the translator. This means I am using trust.

Trust is based on rules and these rules are related to those who give trust and those who receive trust. The bond between those who trust (i.e. the trusters) and those who are trusted (i.e. the trustees) is based on four characteristics of the trustees.

Characteristics of Trustees

What are the characteristics trustees are expected to have?

The first is reputation. One evaluates the trustee's past actions and conduct, and if they are good, then the trustee can be trusted.

The second is performance, which means that one accepts the present actions of the trustee and compares them with what is required to fulfill the responsibilities in question.

The third is confidence, which means that one is pretty sure that expectations of performance will be fulfilled.

And fourth is the most important characteristic of the trustee: competence, which means that the trustee has the knowledge, skills, talents and traits required to be able to perform a task to a given standard.

Trust & Records

If we cannot verify the trustworthiness of digital records on the records themselves, then we have to trust who keeps the records, that is, their custodian. We need to substitute the verification of the authenticity of the record with our trust in the keeper of the records, either the records manager or the archivist.

But of course, this means that we have confidence in the knowledge and ability of the record keeper to perform its tasks in preserving the digital records, based on its reputation.

The level of trust required is proportional to the sensitivity of the material to be trusted and to the consequences of the fact that the material might lose authenticity over time. If we cannot prove that a specific record is authentic, this is not a problem if the record in question is a letter to a friend, but, if the record is a contract for a house, there are big consequences. So the level of trust we need relates to the type of record and/or information and to how important it is to prove that the record is authentic.

In addition, those we trust, records managers or archivists, must be accountable to somebody, must respond for their actions, and operate within a framework of policies, procedures, and technologies.

Types of Digital Records

In order to continue our discussion of the preservation of the trustworthiness of digital records, we have to first look at what kinds of digital records exist. There are basically three categories of digital records. The first category is called computer stored records. They are very much like traditional records, only they are inside a computer: e-mails, reports, any sorts of textual documents, pictures, drawings, etc. This type of records is actually used as evidence of what they say, of their content.

A second category is computer generated records. These are records which are the output of a computer program. There is no human intervention in generating them. For example, when one goes to an ATM to get money, a series of records is generated within the system, as it talks with the bank and checks that one has the money, and records how much money one is taking out of the account, etc. These kinds of records are not used

for their content, but as evidence of the action one has carried out: they attest to the fact that one used the bank ATM and did certain things and so much money was withdrawn.

The third category is a combination of the other two. For example, if one creates a spreadsheet, one includes in the spreadsheet one's own data, which is a human statement. But, the program of the spreadsheet processes the data, so the result is a record that is both stored in and generated by the computer. When one has a record like a spreadsheet, one cannot just preserve the documentary form, but has to preserve its functionality, the way it works in the system.

So, when one preserves computer stored records, it is enough to maintain what one sees on the screen. When one preserves computer generated records, one must preserve the way in which they interact with each other. But when one preserves a spreadsheet, one has to preserve both.

Types of Trustworthiness

Before proceeding to how we do that, we have to think about what trustworthiness means traditionally. Records trustworthiness encompasses three attributes of the records: reliability, accuracy and authenticity.

Reliability is the trustworthiness of the record as a fact. For example, my certificate of citizenship is evidence of my citizenship. So, I trust it as being my citizenship. When we trust the record for what it says, we traditionally accept it at face value without question. We look at who is the author, and, if we trust the author, we trust the record. If the record is complete, if all the parts are there, then we can trust it. If its creation is controlled, then we can trust the record. We can trust the content of the record, because we trust the author, the form, and the process, by inference. For example, if a diagnosis is signed by a doctor, one trusts it, but if it is signed by a nurse, one does not.

Accuracy refers to how correct and precise the data inside the record are, and we base the assessment on the same factors on which we assess reliability, but in addition, on the controls on the way of recording and transmitting the content. For example, if we had a table with columns and rows, when one transmits it, the data might change place. Transmission is the weakest link in the chain of preservation of a record.

Authenticity is the trustworthiness of the record as a record. That means no one has tampered with it, or the record has not been corrupted and has not changed since creation. Authenticity means that a record retains its identity and its integrity.

Reliability

I explained how we used to assess reliability, accuracy and authenticity in the traditional record environment. In the digital environment, if we look at reliability, we can see that the source of the record is still the key. We think the record is reliable if we trust the source. However, with digital records, the source is no longer only a reliable person or a reliable procedure, but can be a process or software: if one trusts the software that generates the record, then one may trust the record.

This implies that the software should be an open-source software, because, if we want to assess the reliability of the record on the basis of a reliable process of creation, then we need to know what that process is; if the software is proprietary, we don't know what it is. So, we need to be able to describe the process, or the system

producing a certain result, or we have to demonstrate the process or the system, and show that it does produce an identical result.

Accuracy

In order to assess accuracy with traditional records, it was enough to demonstrate that the records were original, because one couldn't have changed just the data within the record without the chance of being found out, but with digital entities, as one can change the data without being spotted, one can only demonstrate that the record is accurate if one can repeat the same process of creation and obtain the same result. So, repeatability is one of the fundamental precepts of digital forensics, which is the discipline that has been developed to identify evidence and to prove that evidence has not been forged. The test of accuracy must be supported by the documentation of every action carried out on the record, so one must be able to document everything done to the record.

Of course, in order to repeat a process, one needs to have open-source software, to know what the process was. And this is especially important for archivists, who do conversion and migration of digital records, moving them to different media, and when the media become obsolete, to a different operating system, and so on. Archivists have to be able to prove that, no matter who is doing the process, or under what condition the process is done, the same process will give the same outcome.

Authenticity

When it comes to authenticity, one has to rely on the contexts of the digital records - the procedural context, the documentary context, and the technological context, but also, on the identity and integrity of the records.

The identity of the records used to be provided by the date, the author, the signature, the seal, the classification code, the registry number, etc. Now the identity of the records is in the metadata. So, it is very important that the required metadata are preserved, especially the metadata which show the relationship of the record with the other records, that is the documentary context of the record.

Integrity means that the message that the record is supposed to communicate has not been substantially altered. But what does that imply?

Integrity

With traditional records, when would one say that a record does not have any more integrity? When it has a hole in it? Two holes? When it becomes yellow? When the ink bleeds through? When it is cut? When it is wrinkled? We used common sense to make such decision.

Data Integrity

In the digital environment, the most important thing about integrity is bitwise integrity, which means that the data are not modified either intentionally or accidentally without proper authorization.

Loss of Integrity: Analog vs. Digital

Let's look at this traditional record. We have to assess its integrity.

[Figure 1] How do we know that it has integrity?

[Figure 2] Let's see. It is a bit faded here, but we can still read it.

[Figure 3] Here, it is a little less clear but it still has integrity. If we use a magnifying glass, we can still read it, so we can still say that this is an authentic record because it is still whole there.

Let's consider now a digital record whose original bits are 101. These bits mean 5.

But, if we change their order to 110, that is a six.

If we change it again to 011, then it is a 3.

Same bits, different value. If one changes the order of the bits, even if they are exactly the same, one has a different meaning.

So, integrity is no longer related to degree of deterioration.

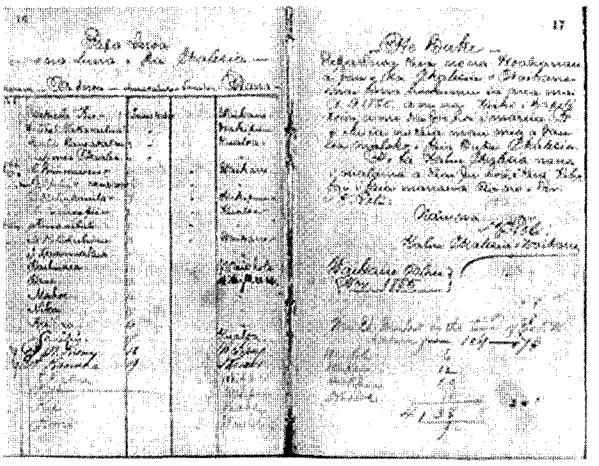


Figure 1

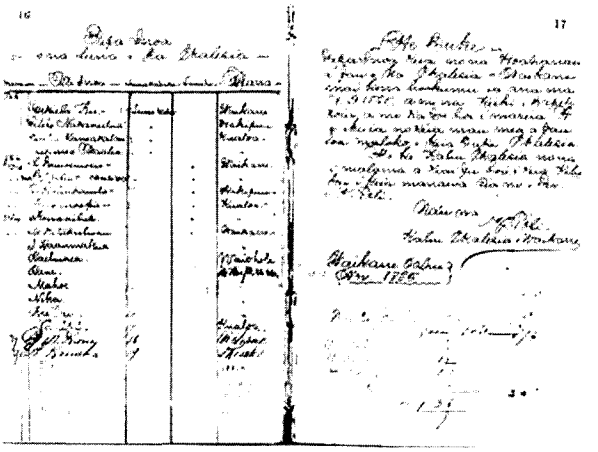


Figure 2

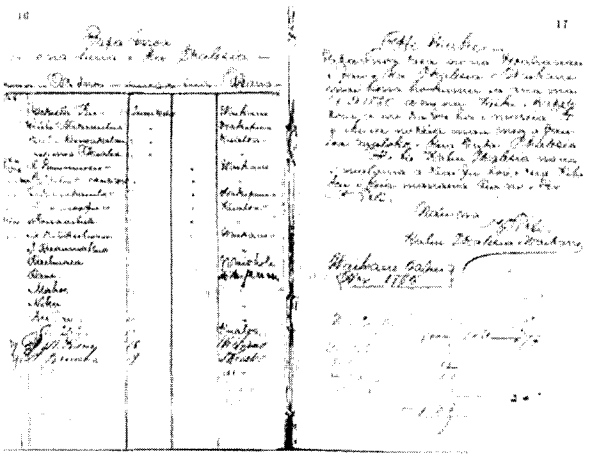


Figure 3

Protecting Records From Data Alteration

Lack of integrity in the digital environment is primarily data alteration, change of data. Alteration can be prevented through permission and access control. We have passwords, firewalls, etc. and strong methods like Checksum and the HASH algorithm. We use a software that basically transforms a record into a string of numbers, and every time one uses the same software with the same record, one must get the same set of numbers. If this does not happen, it means something in the record has been changed.

Accidental alteration is more difficult to avoid than intentional alteration, because it requires a special hardware and software that keeps verifying the records.

In a digital environment, every time one turns on or off the computer, one can create change in the records in the system. And every time one accesses the system, one changes the environment in which the records exist. The police experts who try to sequester records for a

crime, first turn off the computer, then they make an image of the hard drive; they do not try to search anything in the computer, because that can accidentally change the bits. Rather they search the hard drive image.

Prevention is very important, but it is equally important to be able to find out when change has occurred. This is why preservation of digital records is a very laborious effort, because one must continually check to see that change has not happened.

How does one check? By using logs. Logs are files which are automatically created by the system to track all the actions taken by the people who interact with the system.

We will shortly return to logs, but now we have to talk about duplication integrity.

Duplication Integrity

In the traditional environment, when we make copies of records, we are concerned about their accuracy, but we still have the original, so we are not very concerned about the integrity of the copy.

In the digital environment, every time we create a duplicate, that is our record; we no longer have an original. Duplication integrity is defined as the fact that, if we have a record, or a data set, the process of creating the duplicate does not modify the record or data, and the duplicate is an exact digital copy of the original record or data.

In fact, every time we make a copy, what we produce is slightly different from what we had before. This is the reason why it is good to have a time stamp on the copy that one makes: if we have copies taken at different times, we would know that they are different because they were made at different times, not because one is a forgery.

It is important to understand the difference between a copy and an image, because if we tell a forensic expert that we want a copy of an hard drive, what the forensic expert hears is "image of the hard drive," and they are two very different things.

A disk image is a bit-by-bit reproduction of the hard drive. The full disk copy of the data on the hard drive includes also all the empty spaces and all the deleted files.

I know of university archives that take images of all the hard drives of the heads of the departments, because this way, they have the complete and accurate record of what was generated if the computer fails or material is accidentally deleted. However, they are legally on a very shaky ground, because the images also preserve all the files that the heads of the departments had deleted. So, there are ethical problems as well as privacy problems with this procedure.

It is better to make a real copy, which is a selective duplicate of files. One only copies what one can see, not everything that ever was inscribed on the hard drive, because one would need permission to access certain files. So, one should have an incomplete picture of the digital device, and in our preservation responsibilities, we cannot expect to have a complete picture.

On the other hand, after an archives has acquired the records, and it has its own hard drives with all the material, then the best way of keeping reproducing it is to image the hard drive.

Computer and System Integrity

In addition to bitwise and duplication integrity, it is important to consider computer integrity, which means that the computer process produces accurate results when it is used and operated properly, but also that it was so employed when the record was generated.

We also need system integrity. So, you see, we go from the bits up to the record, to the computer, and to the system, because we need all the levels of integrity. We have system integrity when a system performs its intended functions in an unimpaired manner, free from unauthorized manipulation, whether intentional or accidental, and it did so when the record was generated and used.

Computer and system integrity are protected by the usual things: user permissions, passwords, firewalls, but the most important things are the logs mentioned earlier, which are automatically generated by the system.

I am not going to describe all the logs (which are listed on the slide), but I will point out that, of all the logs, most of which can be destroyed after verification, the ones that are important to keep are the auditing logs. The auditing logs are like the black box of a plane, because they will tell who did what, when and where.

Process Integrity

Process integrity is the most important kind of integrity for archivists. In the digital environment, we have to be able to prove what we have done to the records, step-by-step, from the moment we have received them to forever, and what we have to prove is that either we did not interfere with the records, that is, the methods we used to gather them, capture them, use them, or manage and preserve them did not change them, or that if we changed them, we documented the changes.

Authentication

Now a word about authentication is in order. Many legislative texts in many countries, especially in Europe, confuse authenticity with authentication. Legislators think that, if they prescribe a method of authentication, they have guaranteed the authenticity of the records, but that is not true.

Authentication is simply one of the means of declaring that a record is authentic. But, it can only declare that a record is authentic in one specific moment in time, when the declaration is made. Authentication does not keep the record authentic.

The digital signature, for example, has more the function of a seal than the function of a signature, because the signature assigns responsibility for the content of the record and it is a necessary component of the record, while the digital signature is an attachment to a complete record.

The problem with the digital signature is that it cannot be preserved with the record, so it is useful for the transmission of the record, but when one receives it, one cannot preserve it with the record, because it becomes obsolete before the record, and cannot be migrated with the record.

Preferred Means of Authentication

There are other methods for authenticating digital records, and they tend to be procedural methods. A chain of legitimate custody is the best way for inferring authenticity and authenticating a record, because it proves

that the record has been under responsible, trusted custody from the moment it has been generated.

The digital chain of custody is the recording of the information about the record and its changes and shows that specific data was in a particular state at a given time and date. Thus, also this is a good authentication method.

A declaration made by an expert on the trustworthiness of the recordkeeping and the preservation systems is more important than any digital signature.

Preservation

What we have discussed has very important consequences for the meaning of the concept of preservation, and for the function of the archivist.

The concept of preservation in the digital environment must include all the processes necessary to transmit the record through time from creation to forever, including conversion and migration. Preservation is not just keeping what we have, but ensuring that we create records in such a way that we will be able to preserve them. The unbroken chain of preservation must begin with the creation of the record and continue from the record-making system–the system in which the records are generated–to the recordkeeping system, and then to the record preservation system. When we describe a preservation process, we must begin from the moment in which the system where the records are going to be created is designed.

Archivist as Trusted Custodian

The implication is that archivists must present themselves as the trusted custodians of the records.

The trusted custodian is a person who acts as a neutral third party, demonstrates that has no stake in the contents of the records, no reason to alter the records under custody, and that will not allow anybody to alter the records either accidentally or on purpose.

The trusted custodian must have the knowledge and the skills necessary to fulfill his responsibilities, to have competence as the foundation of trust, and should acquire them through formal education.

The trusted custodian must establish a trusted preservation system that is capable of ensuring that accurate and authentic copies of the creator's records are acquired and preserved.

Consider the case of police departments. Traditionally, the police keeps the evidence in the evidence room. However, trials may go on for five years or ten years, and during that time, digital records become obsolete.

This means that they need to be migrated to a new system and a new format to remain accessible and usable at trial. However, when one maintains the records accessibility through these operations, one has the opportunity to change them. So, we cannot have the police, who is one of the two interested parties, migrate the records, because even if they do not change them, we do not trust that this is the case and there is no way of proving it. Thus, the best thing to do is to entrust the digital records to the archivists of the court, because they are a neutral third party. They can do the migration without suspicion that there has been tampering with the records.

The Archivist's New Role

The archivist has acquired a new role – we are really a different profession now.

The archivist now must be situated at the beginning of the record life-cycle, taking the role of the "designated" trusted custodian; must collaborate with the records manager and need to be identified by the record creator as the designated trusted custodian of the records.

The archivist must assess the authenticity of the records over time and monitor their authenticity throughout the existence of the records, because every time the records are moved by the creator from a system to another they might lose their authenticity, their integrity or their metadata.

The archivist must identify the records to be preserved at the moment they are created. Even if this would be a conservative appraisal, it is necessary to identify right away what must be protected so that the archivist will monitor how these records are transformed over time through systems change.

The archivist must determine the feasibility of preservation of the records on the basis of the archives technological capacity. This is very important, because in the past we used to accept whatever was transferred to the archives. If we accept digital material we are unable to preserve, that is equivalent to destroying it. We have to assess our capability of preserving the material before accepting it.

The archivist has to determine a preservation strategy which is based on a controlled process of migration of the records that we have acquired, and remember that, although we have to migrate all the records that we have acquired, we also have to keep the native format because we can never check every record after migration and, if a researcher finds out that there are corrupted records, it is too late if we have not kept the native format.

The archivist has to document all the changes to the record which happen when the technological environment of the archives is upgraded.

Archivists have to designate who in the archives has the right to look at the records, who has the right to carry out actions on the records that can affect the records, who can use the records, who can reproduce the records, in order to avoid that the archivists themselves, when carrying out their functions, alter the integrity of the records.

The archivist has to establish procedures to prevent, discover or correct loss or corruption of records, as well as to guarantee the continuing integrity of the record when there is media deterioration and technological change.

The archivist needs to authenticate individual records according to rules that determine the responsibility for authentication; control the accuracy of the records after each conversion or migration, and develop procedures that address issues of intellectual rights and privacy.

The archivist has to use archival description as a means of authentication of each archival aggregation as a whole, and of the relationship among the records within the aggregation, because, while with traditional records the relationships among the records are evidenced by the place in which the records are, in the digital environment they are not. Only archival description makes them explicit, so the traditional inventory of records is much more important in the digital environment than it ever was in the paper environment, and it functions as an authentication of the documentary context of the record.

Finally, archivists must constantly be involved in research and development projects, just like the industry

does, because we can't wait that scholars, and even less legislators or bureaucrats, tell you what to do. Things change too fast, and archivists must constantly be on the leading edge of research and test of the findings of all research projects that are carried out worldwide, because otherwise, they will never catch up.

InterPARES 3 Products

To start with, you can look at the products list of InterPARES 3. I am not going to describe them, as they can be found on the InterPARES website, under "Products," and then "InterPARES 3." Products can be accessed by case study or by general study, or through the reports of each team, or by keyword or subject matter. Furthermore all InterPARES products can be used at your pleasure, as you wish.

Thank you.

[Notes]

(1) All the key products listed here can be easily accessed through the InterPARES portal at www.interpares.org.

(2) The status of transmission of a record is its degree of perfection. Thus, a draft is incomplete and meant for correction, an original is the first complete record capable of reaching the purposes for which it was intended, and a copy is a reproduction of a draft, an original, or another copy. In the paper world, each record continued to exist in the status of transmission in which it was filed. In the digital world, an original only exists after reception until the moment it is saved. After that we have only copies. A draft only exists as such for as long as one works on it. After it is saved we can only retrieve copies of it.

〔スライド 1〕

InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

TEAM Canada

Trusting Digital Records: the Major Findings of the InterPARES Project"

Luciana Duranti
InterPARES Project Director

Kyoto, Japan 23 June 2012

〔スライド 2〕

The Goal of InterPARES 1 and 2 (1999-2007)

To develop the body of **theory** and **methods** necessary to ensure that digital records produced in **databases** and **office systems** as well as in **dynamic, experiential and interactive systems** in the course of artistic, scientific and e-government activities can be created in **accurate** and **reliable** form and maintained and preserved in **authentic** form, both in the long and the short term, for the use of those who created them and of society at large, regardless of technology obsolescence and media fragility.

〔スライド 3〕

Goal of InterPARES 3 (2007-2012)

To **enable** public and private **archival organizations and programs** with limited resources **to preserve** over the long term **authentic records** that satisfy the requirements of their stakeholders and society's needs for an adequate record of its past.

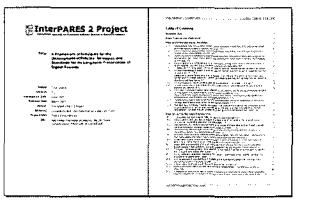
It did so by building on the products of the first two phases of InterPARES (1998-2006)

〔スライド 4〕

Key IP 1 & 2 Products

Policy Framework

A framework of principles guiding the development of policies for records creating and preserving organizations

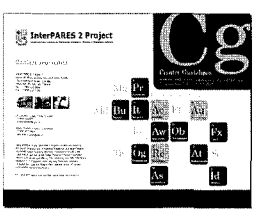


〔スライド 5〕

IP 1 & 2 Products

Creator Guidelines

Recommendations for making and maintaining digital materials for individuals and small communities of practice

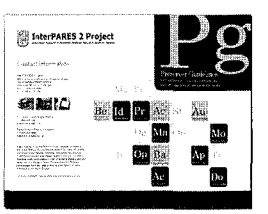


〔スライド 6〕

IP 1 & 2 Products

Preserver Guidelines

Recommendations for digital preservation for archival institutions

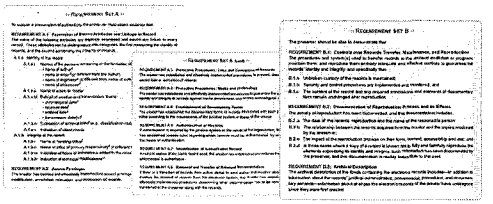


〔スライド 7〕

IP 1 & 2 Products

Benchmark and Baseline Requirements

Authenticity requirements for assessing and maintaining the authenticity of digital records



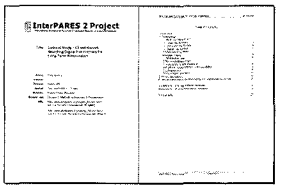
InterPARES Project
Luciana Duranti
Project Director
7

〔スライド 8〕

IP 1 & 2 Products

File Format Selection Guidelines

Principles and criteria for adoption of file formats, wrappers and encoding schemes



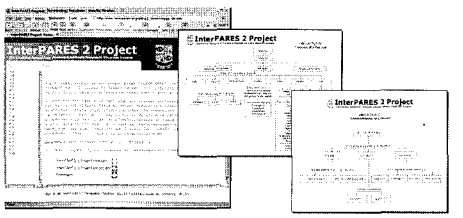
InterPARES Project
Luciana Duranti
Project Director
8

〔スライド 9〕

IP 1 & 2 Products

Terminology Database

Including a glossary, a dictionary and ontologies



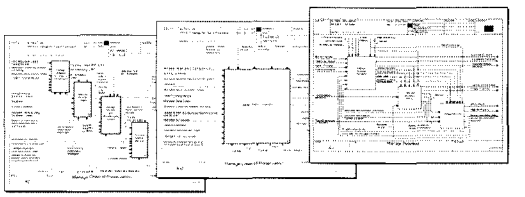
InterPARES Project
Luciana Duranti
Project Director
9

〔スライド 10〕

IP 1 & 2 Products

Two Records Management Models

Chain of Preservation (COP) Model (lifecycle)
Business-driven Recordkeeping (BDR) Model (continuum)



InterPARES Project
Luciana Duranti
Project Director
10

〔スライド 11〕

IP 1 & 2 Final Products

Two books:

Luciana Duranti, ed. *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (San Miniato: Archilab, 2005). Available on line at http://www.interpares.org/book/index.cfm

Luciana Duranti and Randy Preston, eds. *InterPARES 2: Interactive, Dynamic and Experiential Records* (Roma: ANAI, 2008). Available on line at http://www.interpares.org/ip2/book.cfm.

InterPARES Project
Luciana Duranti
Project Director
11

〔スライド 12〕

InterPARES 3 International Alliance

- **Teams:** TEAM (**T**heoretical **E**laboration into **A**rchival **M**anagement) Canada (including US); Brazil; Catalonia; China; Colombia; Italy; Korea; Malaysia; Mexico; Norway (till 2009); Singapore (till 2009); and Turkey.
- **Director:** Luciana Duranti
- **Headquarters:** UBC - SLAIS (facilities provided by UBC)
- **Funding:** SSHRC, and various sources from each country

InterPARES Project
Luciana Duranti
Project Director
12

〔スライド 13〕

InterPARES 3 Findings

- **Conceptual**
  - The Concept of Record
  - The Concept of Trustworthiness
  - The Concept of Life Cycle
- **Methodological**
  - Appraisal
  - Preservation Concept and Procedure
- **Strategic**
  - Relationship Creator-Preserver
  - The Role of the Archivist

InterPARES Project
Luciana Duranti
Project Director
13

〔スライド 14〕

Digital vs. Traditional Records

In the digital environment:

- Record content, structure and medium are no longer inextricably linked
- The stored entity is distinct from its manifestation and its digital presentation has to be considered as well as its documentary one
- When we save a record, we take it apart in its digital components, and when we retrieve it, we reproduce it (it is not possible to preserve a digital record, only the ability to reproduce or recreate it)

Therefore, we can no longer determine trustworthiness on the object-record, which is composite (stored + manifested) and permanently new (re-production), but must **infer trustworthiness from its environment of creation, maintenance & use and preservation.**

InterPARES Project
Luciana Duranti
Project Director
14

〔スライド 15〕

Trust & Its Rules

Trust involves acting without the knowledge needed to act. It consists of substituting the information that one does not have with other information.

The rules of trust refer to those who give trust as well as to those who receive trust:
trusters [givers] and *trustees* [receivers]

The trust-bond between trusters and trustees is usually based on four
characteristics of the trustees

InterPARES Project
Luciana Duranti
Project Director
15

〔スライド 16〕

Characteristics of Trustees

- *reputation*, which results from an evaluation of the trustee's past actions and conduct;
- *performance*, which is the relationship between the trustee's present actions and the conduct required to fulfill his or her current responsibilities as specified by the truster;
- *confidence*, which is an assur-ance of expectation of action and conduct the truster has in the trustee; and
- *compe-tence*, which consists of having the knowledge, skills, talents, and traits required to be able to perform a task to any given standard

Sztompka P (1999) *Trust*. Cambridge University Press, Cambridge

InterPARES Project
Luciana Duranti
Project Director
16

〔スライド 17〕

Trust & Records

- In the digital environment trust in records is an inference based in large measure on *confidence* in the *performance* and *competence* of the keeper of the material, as revealed by its *reputation*.
- The level of trust required is proportional to the sensitivity of the material to be trusted and the adverse consequences of its lack or loss of trustworthiness.
- To guarantee the trustworthiness of **digital records** requires intentional action or intervention by trusted entities imbued with accountability, but also an adequate framework of policies, procedures, and technologies.

InterPARES Project
Luciana Duranti
Project Director
17

〔スライド 18〕

Types of Digital Records

- **Computer Stored Records:** Contain human statements; if created in the course of business, they are records; e.g. e-mail messages, word processing documents, etc. Used as **Substantive Evidence** (of its content)
- **Computer Generated Records**: Do not contain human statements, but are the output of a computer program designed to process input following a defined algorithm; e.g. server log-in records from Internet service providers, ATM records. Used as **Demonstrative Evidence** (of the action from which they result)
- **Computer Stored & Generated:** A combination of the two: e.g. a spreadsheet record that has received human input followed by computer processing (the mathematical operations of the spreadsheet program). Used both or either way.

InterPARES Project
Luciana Duranti
Project Director
18

〔スライド 19〕

Types of Trustworthiness

Reliability: The trustworthiness of a record as a statement of fact, *based on* the competence of its author, its completeness, and the controls on its creation

Accuracy: The correctness and precision of a record's content, *based on* the above, and on the controls on content recording and transmission

Authenticity: The trustworthiness of a record that is what it purports to be, untampered with and uncorrupted, *based on its* identity and integrity, and on the reliability of the records system in which it resides

InterPARES Project
Luciana Duranti
Project Director
19

〔スライド 20〕

Reliability

Reliability: the *source* of the record is the key, defined in a way that points primarily to a reliable person and procedure (for computer stored documents) or a reliable process and software (for computer generated documents), or both.

The software should be open source, because the processes of records creation and maintenance can be authenticated either

- by describing the process or system used to produce a result or
- by showing that the process or system produces an accurate result

InterPARES Project
Luciana Duranti
Project Director
20

〔スライド 21〕

Accuracy

Digital entities are guaranteed accurate if they are repeatable.

Repeatability, which is one of the fundamental precepts of digital forensics, is supported by the documentation of each and every action carried out on the record.

Open source software is again the best choice for assessing accuracy, especially when conversion or migration occurs, because it allows for a practical demonstration that nothing could be altered, lost, planted, or destroyed in the process

InterPARES Project
Luciana Duranti
Project Director
21

〔スライド 22〕

Authenticity

Context: The procedural, documentary and technological environment in which the record was created and used overtime

Identity: The whole of the attributes of a record that characterize it as unique, and that distinguish it from other records (e.g. date, author, addressee, subject, identifier).

Integrity: A record has integrity if the message it is meant to communicate in order to achieve its purpose is unaltered (e.g. text and form fidelity, absence of technical changes).

InterPARES Project
Luciana Duranti
Project Director
22

〔スライド 23〕

Integrity

The quality of being complete and unaltered in all **essential** respects. We were never fussy about it. What if a letter had holes, or was burned on the side or the ink passed through?

The same definition used with respect to data, documents, records, copies, records systems

As long as it was good enough...but how good is good enough in the digital environment?

InterPARES Project
Luciana Duranti
Project Director
23

〔スライド 24〕

Data Integrity

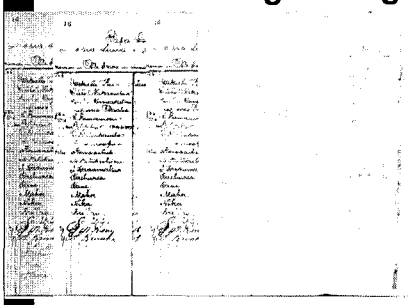
Based on **Bitwise Integrity:** the fact that data are not modified either intentionally or accidentally "without proper authorization."

- The original bits are in a complete and unaltered state from the time of capture, that is, they have the exact and same order and value
- Small change in a bit means a very different value presented on the screen or action taken in a program or database.

InterPARES Project
Luciana Duranti
Project Director
24

〔スライド 25〕

Loss of Integrity: Analog vs. Digital



〔スライド 26〕

Loss of Integrity (cont.)

- If Original Bits 101
- Change state to 110
- Continues to a 011

- Same bits, but
  Different value

3

〔スライド 27〕

Protecting Records From Data Alteration

- Intentional alteration preventable through permission and access controls and strong methods like Checksum and HASH Algorithms
- Accidental alteration avoidance requires that additional hardware and/or software be in place
- We also need methods of determining if the record has been altered, maliciously or otherwise
- Cannot rely on file size, dates or other file properties
- We need logs: sets of files *automatically* created to track the actions taken, services run, or files accessed or modified, at what time, by whom and from where

〔スライド 28〕

Duplication Integrity

The fact that, given a data set, the process of creating a duplicate of the data does not modify the data, and the duplicate is an exact bit copy of the original data set. Time stamps are useful to support it.

Disk Image: a bit by bit reproduction of the storage medium. A full disk copy of the data on a storage device, of the empty spaces and the deleted files

Different from a copy: a selective duplicate of files

- You can only copy what you can see
- Rarely includes confirmation of completeness
- Moved as individual files
- Provides incomplete picture of the digital device

〔スライド 29〕

Computer and System Integrity

Computer integrity: the computer process produces accurate results when used and operated properly and it was so employed when the evidence was generated.

System Integrity: a system performs its intended functions in an unimpaired manner, free from unauthorized manipulation whether intentional or accidental, and it did so when the evidence was generated and used.

Both imply **hardware and software integrity**

〔スライド 30〕

Computer or System Integrity

Protected by:

- Sufficient security measures to prevent unauthorized or untracked access to the computers, networks, devices, or storage.
  - Users/permissions
  - Passwords
  - Firewalls
- **System and Auditing Logs:** Web logs (Client IP Address, Re quest Date/Time, Page Requested, HTTP Code, Bytes Sent, Browser Type, etc.); Access logs (User account ID, User IP address, File Descriptor, Actions taken upon record, Unbind record, Closed connection); Transaction logs (History of actions taken on a system to ensure Atomicity, Consistency, Isolation, Durability; Sequence number; Link to previous log; Transaction ID; Type; Updates, commits, aborts, completes); Auditing Logs (Who-What-Where-When/the black-box)

〔スライド 31〕

Process Integrity

Non-interference: the method used to gather, capture, use, manage and preserve digital data or records does not change the digital entities

Identifiable interference: the method used does alter the entities, but the changes are identifiable

These principles, which embody the ethical and professional stance of records and information managers, archivists, and digital forensics experts, are consistent with the impartial stance of a neutral third party, a trusted custodian

〔スライド 32〕

Authentication

A means of declaring the authenticity of a record at one particular moment in time -- possibly without regard to other evidence of identity and integrity.

Example: the **digital signature**. Functionally equivalent to seals (not to signatures): verifies record's origin (identity); certifies record's intactness (integrity); makes record indisputable and incontestable (non-repudiation). But, seals are associated with a person; digital signatures are associated with a person and a record. They are not a preferred means of authentication through time: they are preferred only across space.

〔スライド 33〕

Preferred Means of Authentication

A chain of legitimate custody is ground for inferring authenticity and authenticate a record.

Digital chain of custody: the information preserved about the record and its changes that shows specific data was in a particular state at a given date and time.

A declaration made by an expert who bases it on the **trustworthiness of the recordkeeping and preservation system** and of the procedures controlling it (**Information governance and quality assurance**).

This has enormous consequences for archival preservation as it affects not only its method but also and primarily its meaning and the role of the archivist.

〔スライド 34〕

Preservation

The traditional **concept of preservation** must include the processes necessary to transmit the record through time, including conversion and migration

The **unbroken chain of preservation** must begin at creation and continue from the record-making system to the recordkeeping system and the record preservation system

The new emphasis on accountability allows the archives to fulfill these needs by **presenting itself as the trusted custodian**

〔スライド 35〕

Archivist as Trusted Custodian

The trusted custodian is a person who

- acts as a **neutral third party**, i.e., demonstrates that he/she has no stake in the content of the records and no reason to alter records under his/her custody, and that he/she will not allow anybody to alter the records either accidentally or on purpose,
- is equipped with the **knowledge and skills** necessary to fulfil its responsibilities, which should be acquired through formal education, and
- establishes a **trusted preservation system** that is capable of ensuring that accurate and authentic copies of the creator's records are acquired and preserved;
- But, mostly...

〔スライド 36〕

The Archivist's New Role

- Positions him/herself at the **beginning of the record life-cycle**, taking the role of "designated" trusted custodian
- Assesses the **authenticity of the records** and **monitors it** throughout their existence
- Identifies the records to be preserved at the moment of their creation and **monitors their transformation through time**
- Determines the **feasibility of preservation** on the basis of the archives technological capacity

〔スライド 37〕

The Archivist's New Role (cont.)

- Determines a **preservation strategy** based on:
  - a controlled process of **migration** of the acquired records to the archives technological environment (always keeping the records also in the format in which they were acquired)
  - the accurate **documentation** of any change that the records undergo during such process and every time that the archives technological environment is upgraded
  - the implementation and **monitoring** of privileges concerning the access, use and reproduction of the records within the archives

InterPARES Project
Luciana Duranti
Project Director
37

〔スライド 38〕

The Archivist's New Role (cont.)

- Establishes **procedures** to prevent, discover, and correct loss or corruption of records, as well as
- Establishes procedures to guarantee the continuing identity and integrity (i.e. **authenticity**) of the records against media deterioration and across technological changes; and
- **Authenticates** individual records according to the rules that determine responsibility for and means of authentication.
- Controls the **accuracy of the records** after each conversion or migration
- Develops **procedures** that address issues of **intellectual rights and privacy**
- Recognizes to **archival description a primary authentication function**
- Is constantly **involved in research and development projects** similar to those carried out by the industry

InterPARES Project
Luciana Duranti
Project Director
38

〔スライド 39〕

InterPARES 3 Products

- Terminology Database
- Directory of Digital Preservation Projects
- Directory of International Standards Relevant to IP3
- E-mail Preservation
- Protocol Registry Preservation
- Community Archives e-Records Assessment
- Public Sector Audit Report for Digital Recordkeeping
- Records Management Policies and Procedures Template

InterPARES Project
Luciana Duranti
Project Director
39

〔スライド 40〕

InterPARES 3 Products (cont.)

- Cost-benefit Models
- Ethical Models
- File Viewers Assessment
- Open Source Records Management Software Assessment
- Metadata Applications Profiles
- Web 2.0/Social Media
- Organizational Culture & Risk Assessment
- Education Modules (with ICA)

InterPARES Project
Luciana Duranti
Project Director
40

〔スライド 41〕

Products on InterPARES Website

- The 3rd phase of InterPARES was completed on 31 March 2012.
- The research findings and products are being uploaded on the public area of the InterPARES website.
- From the home page (www.interpares.org) the material can be accessed by clicking on Products and selecting InterPARES 3.
- Then, you can either click again on Products, and select the material by type (Case Studies, General Studies, Teams Reports), or you can search by keywords, or select the theme you are interested in from the list on the page.
- It will take another couple of months to finalise all products and complete the uploading.

So, keep checking and stay tuned for the next phase.

InterPARES Project
Luciana Duranti
Project Director
41

〔スライド 42〕

Findings and Products

www.interpares.org

InterPARES Project
Luciana Duranti
Project Director
42